# 出柜与否对于性少数群体心理的影响——基于微博知乎数据的研究

王薪舒[12] 赵梦晗[123] 潘超[123] 赵楠[12] 朱廷劭[12]

(1 中国科学院心理研究所，北京，100101)

(2 中国科学院大学心理学系，北京，100049)

(3 河南大学教育科学学院，开封，475004)

**摘要**

**【目的】**探究出柜与否对于性少数群体心理的影响，为性教育以及性少数群体的发展提供建议。 **【方法】**本研究采用 Python 和八爪鱼爬取微博和知乎上性少数群体用户的数据，利用文心系统获得性少数群体情绪词的词频特征，采用机器学习的方法获得生活满意度指标，比较出柜组与未出柜组在词频特征和生活满意度指标上的差异。**【结果】**未出柜组在负性情绪词（$t$ ＝ -3.043, $p$ ＜ 0.01）和悲伤词（$t$ ＝ -2.211, $p$ ＜ 0.05）上的比例显著高于出柜组；生活满意度显著低于出柜组（$t$ ＝ 5.078, $p$ ＜ 0.001）。**【局限】**本研究未出柜组采样不足，未通过年龄进行被试筛选，选择变量较少及横向研究，使得研究不够全面。**【结论】**出柜可以提高性少数群体的生活满意度并促进性少数群体心理健康发展。

**关键词：**性少数群体，出柜，词频分析，生活满意度，机器学习

# Sexual Minorities' Psychology Influence by Coming Out or Not——A Research Based on Weibo, Zhihu Data

Wang Xinshu[12] Zhao Menghan[123] Pan Chao[123] Zhao Nan[12] Zhu Tingshao[12]

1 (Institute of Psychology, Chinese Academy of Sciences, Beijing 100101)

2 (University of Chinese Academy of Sciences, Beijing 10049)

3 (Henan University, Kaifeng 475004)

**Abstract:**

**[Objective]** To explore the impact of coming out or not on the psychology of sexual minorities, and provide recommendations for sex education and the development of sexual minorities. **[Methods]** This study used Python and Octopus to crawl the data of Weibo users and Zhihu users, TextMind to obtain the word frequence characteristics of emotional words, and machine learning methods to obtain life satisfaction indicators, and compared the difference between the word frequency characteristics and life satisfaction indicators of the coming-out group and the non-coming-out group. **[Results]** The proportion of negative emotion words ($t$ ＝ -3.043, $p$ ＜ 0.01) and sad

words ($t$ = -2.211, $p < 0.05$) in the group of non-coming-out group was significantly higher than that of the coming-out group. The life satisfaction of the people who did not come out was significantly lower than that of the people who came out ($t$ = 5.078, $p < 0.001$).
**[Limitations]** The non-coming-out group in this study was insufficiently sampled, and the subjects were not screened by age, selected variables were few and horizontal research, which made the research not comprehensive enough. **[Conclusion]** Coming out can improve the life satisfaction of sexual minorities and promote the mental health of them.

**Key words:** Sexual minorities; coming out; word frequency analysis; life satisfaction; machine learning

## 1 引言

性少数群体主要包括女同性恋、男同性恋、双性恋等，通常指性取向不完全属于异性恋的人 (张静, 郑丽军, 郑涌, 2015; Gonzales & Blashill, 2020; Horwitz et al., 2020)。近几年随着我国当前社会的不断开放与发展，特别是互联网的普及化，各个年龄段、各个地域、各经济地位的人群都能更多地接触到与性及性倾向相关的信息，与此同时，性少数群体数量也在不断增长 (潘绥铭, 2016)。中国是个极具传统家庭观念和伦理道德观的国家，性少数群体的性倾向往往难以被承认和接受(张静等, 2015)。也就是说，在我国这样的大环境下，相比于异性恋群体，性少数群体会受到主流文化的排斥，承受着更多暴力、偏见、歧视等慢性压力源，这些也使得他们存在更多的心理健康问题 (王恒杰, 唐日新, 2019; 张静等, 2015)。目前许多国内研究者已发现我国性少数群体心理健康水平低于国内常模（许毅, 严乐勤, 陈昭典, 2000; 李艳芳等, 2020; 韦丽, 2015）。

生活满意度指对个人整体生活质量或对家庭、朋友和社区等生活特定方面的认知评价 (Hu, Hu, Huang, & Zheng, 2016)，可更全面和客观地反映性少数群体的心理健康水平 (李放等, 2019)。国内已有部分研究证明男女同性恋的生活满意度较低 (田宝伟, 胡心怡, 2016)，但未探讨双性恋群体的生活满意度水平，这就亟需我们扩大研究范围，将双性恋群体纳入其中，进一步探讨性少数群体的生活满意度。

出柜是指性少数群体向他人公开自己的性取向的行为 (张静等, 2015; 张严文, 叶宝娟, 2019)，通常被理解为性少数群体身份发展的重要组成部分，对于他们发展真实和稳定的自我感，培养积极的性少数群体身份，具有积极作用，被视为性少数群体理想的最终状态 (Cain, 1991; Ragins, 2004; Wells & Kline, 1987)。然而目前出柜与生活满意度的关系仍未达成一致。一方面，由于家庭和社会系统假定个体属于异性恋，所以出柜这件事情对于性少数群体而言本身就是有压力的 (Bogaert & Hafer, 2009)。同时，他们在出柜后很有可能会遭受更多的伤害，这也会对他们的生活满意度造成负面影响 (D’Augelli, Pilkington, & Hershberger, 2002; Pilkington & D’Augelli, 1995)。另一方面，Meyer (2003) 认为，出柜可以减少焦

虑、促进自我同化，从而改善自身特质，对自尊和心理健康有着积极的影响 (张静等, 2015)。因此，本研究试图探索出柜对性少数群体生活满意度的影响性质。此外，情绪词频如正向情绪词、负向情绪词、焦虑词、生气词、悲伤词的比例可以客观反映网络用户的心理体验，对于分析性少数群体的心理健康也具有重要意义。因此，本研究也试图探究出柜对于性少数群体情绪词词频的影响。

以往的研究往往使用问卷法和访谈法，被试很可能因为身份的敏感性而拒绝作答或是隐蔽一些真实信息。网络最大的特点便是匿名性，根据去抑制性理论，在匿名情况下人们会倾向于说或者做在现实人际交往中不敢说或不敢做的事，比如性少数群体为获取主流文化的支持、增强少数群体内部凝聚力，会主动在网络上发布一些与自己性取向或者出柜过程相关的事情 (余文斌, 2013)，这为我们分析他们的心理提供了很好的数据来源。知乎和微博都具备一定的匿名性和时效性，性少数群体也往往愿意在知乎、微博上为自己发声，因此本研究采用 Python 和八爪鱼爬取微博和知乎上性少数群体的数据进行分析，增加研究的生态效度。

# 2 方法

## 2.1 被试选择及数据获取

本研究选取性少数群体网络用户，所使用的数据来源于知乎用户的回复和微博活跃用户的微博数据，通过选定的关键词对知乎和微博内容进行检索及爬取。最终获得 3025 条微博数据和 1611 条知乎回复数据，共计 4636 条。

## 2.2 网络爬虫工具及关键词

通过浏览相关文献和网站，人工筛选出一些具有代表性的性少数群体词汇、出柜词汇、未出柜词汇和出柜对象词汇，用来作为数据获取的检索依据，如表 1 所示。随后我们使用 python 及八爪鱼采集器等工具根据不同关键词的组合对知乎和微博两个软件进行爬取。

表 1 检索词汇表

| 性少数群体词汇 | 出柜词汇 | 未出柜词汇 | 出柜对象词汇 |
| --- | --- | --- | --- |
| 同性恋、gay、les、男同、女同、同志、拉拉、蕾丝、双、双性等 | 出柜、坦白…是同性恋、暴露…是同性恋等 | 不敢、害怕、纠结、难以、犹豫 | 父母、朋友、亲戚、爸妈、身边人、周围人、同学等 |

其中，已有研究表明，网络出柜与圈子内出柜不属于出柜 (张严文, 叶宝娟, 2019)，因此，出柜对象没有选择微博或者性少数群体伙伴。

## 2.3 数据清洗

在爬取全部数据后，我们人工对数据进行筛选来清洗出合格数据。具体清洗的标准包括：（1）有人重复在不同问题下进行相同回答，或者发布相同内容微博；（2）有些数据长度过

短；（3）讲述他人故事或无法判断用户自身出柜情况及是否属于 LGB 群体；（4）属于被动出柜，不属于主观出柜的个体的描述；（5）语义中透露当前未出柜但未来想要出柜。将符合上述标准的条目剔除后，知乎可用数据共计 567 条，微博可用数据 35 条，其中进行人工分类出出柜用户数据 537 条，未出柜 65 条。

### 2.3 内容特征提取

“文心”中文心理分析系统 (TextMind) 由中国科学院心理研究所计算网络心理实验室所研发，它包括中文分词工具和心理分析词典。本研究利用“文心”中文心理分析系统 (Text Mind) 对清洗后的数据进行内容特征的提取。

### 2.4 生活满意度预测模型

生活满意度反映了个体对自身状态的主观满意程度，本研究使用机器学习算法建立生活满意度预测模型，将清洗后数据导入该模型后得到了每位用户的生活满意度指标。

## 3 结果

### 3.1 词频分析

本研究利用文心系统中的简体中文 LIWC 词典计算出出柜组与未出柜的情绪词词频比例，利用独立样本 t 检验比较两组样本在情绪词比例上的差异，结果如表 2 所示，未出柜组在负性情绪词 ($t$ = -3.043, $p$ < 0.01 )和悲伤词 ($t$ = -2.211, $p$ < 0.05 )上的比例显著高于出柜组。尽管出柜组和未出柜组在正向情绪词、焦虑、生气上不存在显著差异，但可以由表 2 可以看出，出柜组在正向情绪词的词频高于未出柜组，在生气词的词频低于未出柜组。

表 2 词频分析描述性数据与独立样 $t$ 检验

| 情绪词 | 分组 | $M \pm SD$ | $t$ | $df$ | $p$ |
|---|---|---|---|---|---|
| 正向情绪词 | 出柜组 | 0.025 ± 0.001 | 0.606 | 600.000 | 0.545 |
| | 未出柜组 | 0.022 ± 0.003 | | | |
| 负向情绪词 | 出柜组 | 0.012 ± 0.001 | -3.043 | 75.606 | 0.003$^{**}$ |
| | 未出柜组 | 0.014 ± 0.002 | | | |
| 焦虑 | 出柜组 | 0.007 ± 0.000 | -1.843 | 600.000 | 0.066 |
| | 未出柜组 | 0.007 ± 0.001 | | | |
| 生气 | 出柜组 | 0.004 ± 0.000 | -1.105 | 72.562 | 0.273 |
| | 未出柜组 | 0.006 ± 0.001 | | | |
| 悲伤 | 出柜组 | 0.004 ± 0.000 | -2.211 | 71.490 | 0.030$^{*}$ |
| | 未出柜组 | 0.006 ± 0.001 | | | |

注：$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

### 3.2 生活满意度指标

对机器学习模型处理后的生活满意度数据进行独立样本 t 检验，结果发现，未出柜人群

的生活满意度显著低于出柜人群的生活满意度（$t = 5.078, p < 0.001$），如表 3 所示。

表 3 描述性数据与独立样 *t* 检验

| 生活满意度 | $M \pm SD$ | $t$ | $df$ | $p$ |
|---|---|---|---|---|
| 出柜组 | 0.572 ± 0.068 | 5.078 | 600.000 | 0.000*** |
| 未出柜组 | 0.527 ± 0.062 | | | |

注：$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

## 4 讨论

为了探究出柜与否对于性少数群体心理的影响，采用前人文献筛选出性少数群体词汇、出柜词汇、未出柜词汇以及出柜对象词汇，使用 python 以及八爪鱼采集器在微博、知乎两个头部社交软件进行数据爬取。对爬取的数据进行人工筛选，在原有的 4636 条数据上筛选出 602 条有效数据。根据每条文本的内容将所有数据归类到“出柜组”和“未出柜组”。其中，出柜组共有 537 条数据，未出柜组共有 65 条数据，利用文心系统和生活满意度预测模型获得词频特征及生活满意度指标并进行组间比较，结果发现在性少数群体中，自主选择出柜的群体的生活满意度显著高于未出柜群体，出柜群体在负性情绪词和悲伤词的比例显著低于未出柜组。

未出柜群体生活满意度显著低于出柜群体，研究结果与以往大部分的研究一致。性少数群体往往由于知觉社会压力和社会歧视而产生内化的对自己性取向的厌恶和抵触，因此他们不会主动接触性少数群体，也不会主动暴露自己的性取向，很难具备良好的身份认同感。这种性取向隐瞒的行为会严重损耗个体的心理，导致更少的积极情绪和生活满意度资源 (Hu, Wang & Wu, 2013)。词频结果中未出柜组在负性情绪词和悲伤词上的比例显著高于出柜组，此外，尽管不存在显著差异，但也发现出柜组在正向情绪词的词频高于未出柜组，在生气词的词频低于未出柜组。这都证明了出柜有助于心理健康水平的提升 (张静等, 2015; Meyer, 2003)。

本研究存在一定的局限性：首先，部分性少数群体更倾向于隐藏自己的性取向，因此，对这一部分群体采样不足，这与以往的研究和预期的样本量并不一致。究其原因，可能以往的研究采用的是滚雪球的抽样方法，是由几个少数群体被试再提供另外的少数群体被试，存在信任的问题。但是在网络上，决心深柜的群体主动公开的情况还是很少。未来的研究应专注于选择特定的方法来挖掘倾向于隐瞒的性少数群体。其次，在爬取数据时没有年龄标准可供筛选，尽管国外已经开始研究 12-14 岁左右的性少数群体，但研究者认为性取向一般在 16 岁以后趋于成熟，未来可能需要限制年龄来进行被试的筛选。再次，知乎数据大部分是知乎

用户对于一些问题的回答，筛选后的文本也大都是描述他们出柜的经历，因此，出柜经历是否良好会影响文本中积极词汇和消极词汇的频率，可能会导致研究结果出现偏差。未来可能需要爬取出柜前后一个月内的文本进行差异检验。最后，可能存在某些因素导致出柜与否对于性少数群体生活满意度的影响不同，比如婚姻压力、社会支持、性别等等，未来可加入其他变量进行探究。

除了上述的局限性外，本研究采取 Python 和八爪鱼分别爬取微博用户和知乎用户数据进行分析，探讨性少数群体出柜和心理的关系，存在一定的理论意义和实践意义：首先，性少数群体作为一个特殊群体，我国对于他们生活满意度的研究本身较少，本研究丰富了性少数群体的生活满意度理论研究。其次，以往针对性少数群体的研究通常都采用问卷法和访谈法，他们极可能因为性取向的敏感性而选择隐蔽一些真实信息，而网络数据具有实时性、真实性和开放性，能够相对真实地反映他们的内心世界，具有生态效度高、数据量大等优势。最后，随着性少数群体数量的不断增多，他们的身心健康问题也日益浮出水面，但如何从个体角度帮助其提高心理健康水平仍有待研究。因此，在本研究中，自主出柜对于性少数群体的生活、心理水平有积极作用，那么在未来可以适当的鼓励性少数群体出柜，帮助他们缓解心理压力，提高心理健康水平。



## 6 结论

本研究利用 python 和八爪鱼采集器，使用关键词识别出微博和知乎上的性少数群体，采集生态化数据并通过文心系统和生活满意度预测模型获得性少数群体词频特征与生活满意度指标并将其进行分析，结果发现性少数群体中自主选择出柜的群体的生活满意度显著高于未出柜群体，出柜组在负面情绪词和悲伤词的词频比例上显著低于非出柜组。

本研究尚且存在一些局限，利用微博和知乎性少数群体用户数据，对于更深层次的性取向隐瞒识别力不够，对于当前的研究结果可能会产生一定影响。

综上，随着社会多元文化之间的碰撞以及社会观念的逐渐开放，性少数群体的心理健康也逐渐受到社会的关注，研究显示自主选择出柜的群体表现出更好的心理健康水平，对性教育以及性少数群体的发展提供建议。

## 7 参考文献


李放，黄李佳，钟林鹏，王一博，邢锦涛，郑雪. (2019). 男同性恋者恋爱状况与生活满意度的关系：一个链式中介模型. *中国临床心理学杂志, 27*(4), 785-789.

李艳芳，赖志胜，卢次勇，黄丽娟，湛柳华，戴丽萍，蔡梅英，陈昱丰，陈力. (2020). 男同/双性恋者抑郁状况及儿童期虐待对其影响的研究. *中国艾滋病性病, 26*(11), 1206–1208+1225.

潘绥铭. (2016). 同性之间发生了什么？取自 http://blog.sina.com.cn/s/blog_4dd47e5a0102w7zd.html[2021-11-24]

佘文斌. (2013). 网络社区中的同性恋身份认同. *安徽大学学报(哲学社会科学版), 37*(1), 144-148.

田宝伟，胡心怡. (2016). 压力知觉、歧视知觉及社会支持对同性恋男大学生心理健康的影响. *中国特殊教育,* (12), 91-96.

王恒杰, 唐日新. (2019). 中国同性恋人群心理健康研究综述. *校园心理,* (3), 201–204.
韦丽. (2015). *山东省女同性恋人群的心理健康状况及行为特征研究*. (硕士论文). 山东大学：山东.
许毅, 严乐勤, 陈昭典. (2000). 男性同性恋明尼苏达多项人格调查表测试结果分析. *中国心理卫生杂志, 14*(4), 244.
张静, 郑丽军, 郑涌. (2015). 性少数人群的心理健康:理论模型与研究取向. *心理科学进展, 23*(6), 1021–1030.
张严文, 叶宝娟. (2019). 父母拒绝教养方式对中国同性恋者自杀的影响:歧视知觉的中介效应与出柜的调节效应. *心理科学,* (1), 7.
Bogaert, A. F., & Hafer, C. L. (2009). Predicting the timing of coming out in gay and bisexual men from world beliefs, physical attractiveness, and childhood gender identity/role. *Journal of Applied Social Psychology, 39*(8), 1991-2019.
Cain, R. (1991). Stigma management and gay identity development. *Social Work, 36*(1), 67-73.
D'Augelli, A. R., Pilkington, N. W., & Hershberger, S. L. (2002). Incidence and mental health impact of sexual orientation victimization of lesbian, gay, and bisexual youths in high school. *School Psychology Quarterly, 17*(2), 148-167.
Gonzales, M., & Blashill, A. J. (2020). Ethnic/racial and gender differences in body image disorders among a diverse sample of sexual minority U.S. adults. *Body Image, 36*, 64–73.
Horwitz, A. G., Grupp-Phelan, J., Brent, D., Barney, B. J., Casper, T. C., Berona, J., Chernick, L. S., Shenoi, R., Cwik, M., & King, C. A. (2020). Risk and protective factors for suicide among sexual minority youth seeking emergency medical services. *Journal of affective disorders, 279*, 274–281.
Hu, X., Wang, Y., Wu, C. (2013) Acceptance concern and life satisfaction for Chinese LGBs: The mediating role of self-concealment. *Social Indicators Research, 114*(2), 687-701
Hu, J. C., Hu, J. Z., Huang, G., & Zheng, X. F. (2016). Life satisfaction, self-esteem, and loneliness among LGB adults and heterosexual adults in China. *Journal of homosexuality, 63*(1), 72-86.
Meyer, I. H. (2003). Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: Conceptual issues and research evidence. *Psychological Bulletin, 129*(5), 674–697
Pilkington, N. W., & D'Augelli, A. R. (1995). Victimization of lesbian, gay, and bisexual youth in community settings. *Journal of Community Psychology, 23*(1), 34 –56.
Ragins, B. R. (2004). Sexual orientation in the work place. Research in Personnel and Human Relations, 23, 35-120.
Wells, J. W., & Kline, W. B. (1987). Self-disclosure of homosexual orientation. *Journal of Social Psychology, 127*(2), 191-197.
Yu, Y. & Xiao, S.Y. (2017). Health and life satisfaction for Chinese gay men in Guangzhou, China.. Journal of Central South University. *Medical sciences, 42*(12), 1407-1416.

作者贡献声明：

王薪舒：提出研究思路，设计研究方法，微博数据爬取、清洗与整理，数据分析，论文前言撰写与最终文本格式修订

赵梦晗：提出研究思路，设计研究方法，知乎数据爬取、清洗与整理，论文讨论、结论撰写与文本修改

潘超：提出研究思路，设计研究方法，知乎数据爬取、清洗与整理，数据分析，论文结果撰写与格式修改

赵楠、朱廷劭：技术分析指导、论文指导